A Score-Based Opportunistic Scheduler for Fading Radio Channels

T. Bonald

France Telecom R&D
e-mail: thomas.bonald@francetelecom.com

***Abstract*: While fading effects have long been combatted in 2G wireless networks, primarily devoted to voice calls, they are now seen as an opportunity to increase the capacity of 3G networks that incorporate data traffic. The packet delay tolerance of data applications, such as file transfers and Web browsing for instance, allows the system flexibility in scheduling a user's packets. Opportunistic scheduling ensures transmission occurs when radio conditions are most favorable. This paper discusses different resource sharing strategies and presents some shortcomings of the classical Proportional Fair opportunistic scheduler. A new algorithm, called the Score-Based scheduler, is presented and shown to overcome these drawbacks.**

***Keywords:* Fast fading, multi-user diversity, opportunistic scheduling.**

1 Introduction

Data services are expected to constitute a significant part of traffic in 3G wireless networks. A number of new technologies have recently been standardized to support high data rates and optimize the spectrum utilization of downlink channels. High Data Rate (HDR) systems [5], defined in the 3GPP2 cdma2000 IS-856 standard [3, 12], offer a maximum data rate of 2.4 Mbit/s over a signal bandwidth of 1.2 MHz, while their 3GPP equivalent High Speed Downlink Packet Access (HSDPA) systems [2, 1, 19] offer a maximum data rate of around 10 Mbit/s over a signal bandwidth of 5 MHz. These systems deliver high spectral efficiency by using a TDMA-like strategy (the base station (BS) transmits at full power to only one user in each slot) with a combination of link adaptation, hybrid ARQ and opportunistic scheduling:

- Link adaptation refers to the adaptation of a user's transmission data rate to its radio conditions based on Data Rate Control (DRC) signals sent back by the user to the BS.

- Hybrid ARQ allows the transmission of any packet spread over multiple slots to be terminated early, i.e., as soon as the packet is successfully received, so as to adapt the transmission rate to the *actual* radio conditions. This control scheme, based on Chase combining or incremental redundancy, is essential given the errors in channel quality prediction and the necessarily conservative SINR thresholds used to ensure a successful transmission.

- Opportunistic scheduling seeks to transmit a user's packets in slots when conditions are relatively favorable, based on DRC feedback signals.

These dynamic schemes take advantage of the inherent "elasticity" of data transfers to increase the overall system capacity: instead of wasting radio resources in providing a constant data rate to each user, they dynamically adapt the data rate of each user to optimize the spectrum utilization. The duration of a slot (1.67 ms for HDR systems, 0.67 ms for HSDPA systems) is sufficiently short to benefit from the uncorrelated fast variations of channel quality experienced by the active users. Thus fading effects, which have long been combatted in 2G wireless networks, are now seen as an opportunity to increase the capacity of 3G wireless networks [23].

The scheduling algorithm is a key component of these time-shared systems. In addition to exploiting multiuser diversity over short time-scales, this algorithm also determines how resources are shared over longer timescales. An algorithm that always selects the user with the highest data rate is efficient in term of overall throughput but may starve low SINR users, typically located far from the BS. An algorithm that equalizes the data rates of active users, on the other hand, is fair but inefficient as most radio resources are used to sustain the data rate of distant users [6]. A third strategy, which realizes a reasonable trade-off between efficiency and fairness, consists in transmitting to the user with the highest data rate relative to its current mean data rate [11, 23]. This scheduler, termed Proportional Fair (PF), has been studied in [14, 15, 17] and is widely used in currently developed systems. Many other scheduling algorithms have been proposed and analyzed (see e.g. [4, 9, 10, 18, 20, 22] and references therein).

To appropriately define long-term sharing objectives, it is actually necessary to account for the fact that the actual set of active users is *dynamic* and varies as a random process as new data flows are initiated and others complete. As each flow is characterized by some amount of data to be transferred, the resource attributed to any user determines how long that user will stay active. In particular, an "efficient" strategy that selects always near users results in a steady state where most active users are far from the BS. A "fair" strategy, on the other hand, results in a much more favorable steady state where users are more uniformly distributed in the cell. Thus the efficiency-fairness trade-off becomes largely irrelevant when accounting for flow level dynamics. The fair strategies are often the most efficient in terms of average performance, as already observed in the context of wired networks [7].

In this paper, we first demonstrate that a strategy that fairly shares the radio resource (here the slot) is indeed efficient in terms of average user performance (the mean flow duration). The performance of the classical PF scheduler is then evaluated both in the ideal case considered in [14] and in more realistic cases with asymmetric fading statistics and data rate constraints. It is shown that, while fair and indeed opportunistic in the ideal case, the PF scheduler may be unfair and unable to fully exploit multi-user diversity in the realistic cases. Finally, we present the principle and the key properties of a new algorithm referred to as the Score-Based (SB) scheduler. This scheduler behaves like the PF scheduler in the ideal case but its performance does not suffer from asymmetric fading statistics or data rate constraints.

The next section is devoted to the flow level analysis of different sharing objectives. The principle and the key characteristics of the PF scheduler and the SB scheduler are presented in the following two sections. Their performance is compared by simulation in Section 5. Section 6 concludes the paper.

2 Sharing objectives

In this section, we analyse the impact of different long-term sharing objectives on user performance. Here we do not consider the scheduling gains due to multi-user diversity over short time-scales. Specifically, we neglect user mobility and fading effects and assume the transmission rate of each user (when scheduled) is approximately constant during data transfer and is mainly determined by its location in the cell. We first present the traffic model then successively consider the three above mentioned strategies: fair time slot sharing, maximum rate sharing and fair rate sharing.

A flow level model. Let n be the number of active users and ϕ_i the fraction of slots allocated by the base station (BS) to user i, with $\sum_{i=1}^n \phi_i = 1$. The data rate of user i is then:

$$\phi_i \times R_i, \tag{1}$$

where R_i denotes the transmission rate of user i when scheduled. We assume the slot is infinitely small so that each flow can be represented as a "fluid" data transfer with a rate given by (1) which varies as new flows are initiated and other complete. The system then behaves like a time-shared queue where the server corresponds to the radio resource and the service required by user i (in time slots) is equal to the ratio of the size of its flow σ_i (in bits) to its transmission rate when scheduled R_i. The slot allocation $\{\phi_i\}_{i=1}^n$ defines the service discipline.

We assume data flows arrive as a Poisson process of intensity $\lambda(x)dx$ in any area of surface dx around cell position x. Denoting by σ the mean flow size, $\lambda(x)\sigma$ corresponds to the traffic density in position x (in Kbit/s per surface area). Let $R(x)$ be the transmission rate of a user in position x when scheduled. The load generated in any area of surface dx around position x is defined as the ratio of the traffic intensity $\lambda(x)\sigma dx$ to the transmission rate $R(x)$. This corresponds to the demand on the radio resource (the time slot).

We deduce the overall cell load:

$$\rho = \int_{\text{cell}} \frac{\lambda(x)\sigma}{R(x)} dx.$$

Note that most load is concentrated in those regions of the cell where the transmission rate $R(x)$ is low. In the following, we always assume that $\rho < 1$. This corresponds to the stability condition of the above defined time-shared queue, provided the service discipline is work-conserving (i.e., all slots are allocated when there is at least one active user).

Users experience quality of service through the duration of data flows. We define the mean flow throughput γ as the ratio of the mean flow size σ to the mean flow duration. Applying Little's law, we get:

$$\gamma = \int_{\text{cell}} \lambda(x)dx \frac{\sigma}{\text{E}[n]}. \tag{2}$$

Thus the mean flow throughput is inversely proportional to the mean number of active users in steady state.

Fair time slot sharing. We first consider a strategy that gives the same fraction of slots to each user, i.e.,

$$\phi_i = \frac{1}{n} \quad \forall i = 1, \ldots, n.$$

This allocation may be realized at slot level by a simple round-robin scheduler for instance. The corresponding time-shared system is a *processor sharing queue* [8, 9]. The steady-state distribution π of n is then insensitive to the flow size distribution and given by:

$$\pi(n) = \rho^n(1 - \rho). \tag{3}$$

In particular, we have:

$$\text{E}[n] = \frac{\rho}{1 - \rho}. \tag{4}$$

It follows from (2) that:

$$\gamma = \bar{R}(1 - \rho), \tag{5}$$

where $\bar{R}$ corresponds to the mean flow throughput when $\rho \to 0$, given by:

$$\bar{R} = \left(\int_{\text{cell}} \lambda(x)dx \right) / \left(\int_{\text{cell}} \frac{\lambda(x)}{R(x)} dx \right).$$

Thus user performance depends on traffic characteristics through two parameters only, the mean rate $\bar{R}$ and the cell load ρ. This insensitivity property allows the development of simple dimensioning rules, robust with respect to evolutions in the nature of data applications (refer to [8]).

Maximum rate sharing. We now consider the strategy that selects always the user with the highest transmission rate, i.e.,

$$\phi_i = 1 \quad \text{for } i = \arg \max_{j=1,\ldots,n} R_j.$$

The corresponding time-shared system is a *priority queue*. The steady-state distribution π of n is then highly sensitive to the flow size distribution. In particular, the expected number of active users is infinite if flow sizes have a heavy-tailed distribution, which is indeed characteristic of data traffic. This is due to the fact that the mean duration of starvation periods for low SINR users is infinite (refer to [7] for a similar result in the context of IP networks). We deduce that the mean flow throughput is equal to zero in this case. Thus a strategy that looks optimal in terms of overall throughput in a static scenario proves extremely inefficient in terms of average user performance in a dynamic scenario. This underscores the importance of dynamic flow level analysis when defining sharing objectives.

Fair rate sharing. Finally, we consider a strategy that is fair from a user perspective in the sense that it equalizes the instantaneous data rates. In view of (1), this is realized by allocation a fraction of the slots inversely proportional to the transmission rate, i.e.,

$$\phi_i = \frac{1/R_i}{\sum_{j=1}^{n} 1/R_j}.$$

The corresponding time-shared system is a *discriminatory processor sharing queue*. Again, the steady-state distribution π of n is sensitive to the flow size distribution. For an exponential flow size distribution, the mean number of active flows can be derived from [13] assuming the set of transmission rates is discrete. Specifically, we assume an arbitrary number of K feasible transmission rates $R_1, \ldots, R_K$ is available. We denote by ρ_k the load of the region of the cell where users have the transmission rate R_k, with $\rho = \sum_{k=1}^{K} \rho_k$. It follows from [13] that:

$$\mathrm{E}[n] = \frac{\rho}{1-\rho} + \frac{\left(\sum_{k=1}^{K} \rho_k R_k\right)\left(\sum_{k=1}^{K} \rho_k / R_k\right) - \rho^2}{(1-\rho)(2-\rho)}.$$

Observing that the second term of last expression is non-negative, we deduce from (4) that the mean number of active users is always larger than that obtained with fair time slot sharing. In view of (2), we conclude that the mean flow throughput is always worsened by a fair rate sharing strategy.

3 Proportional Fair scheduler

In the rest of the paper, we analyse the behavior of the PF scheduler and the SB scheduler over short time-scales. Specifically, we assume that the number of active users n is fixed and study the long-term sharing realized by these schedulers and their ability to exploit multi-user diversity. In view of the results of Section 2, the scheduler should share the time slots in a fair way to ensure good performance at the flow level, insensitive to the flow size distribution.

This section is devoted to the PF scheduler. We first recall the principle of the algorithm then evaluate its performance in various conditions.

We denote by $r_i(t)$ the transmission rate of user i at slot t if scheduled. We assume $r_i(t)$ is a stationary and ergodic process. Let R_i be the mean transmission rate of user i, equivalent to its realized throughput in the absence of any other user:

$$R_i = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} r_i(t).$$

We denote by $\xi_i(t) = r_i(t)/R_i$ the normalized rate.

Principle. The PF scheduler selects at slot t the user $i(t)$ with the highest transmission rate relative to its current average throughput, i.e.,

$$i(t) = \arg \max_{j=1,\ldots,n} \frac{r_j(t)}{T_j(t)}, \tag{6}$$

where the throughput $T_j(t)$ is typically evaluated through an exponentially smoothed average:

$$T_j(t) = (1 - \frac{1}{t_c}) \times T_j(t-1) + \frac{1}{t_c} \times r_j(t-1) \mathbb{I}_{\{i(t-1)=j\}}.$$

The time constant t_c captures the time-scale of the PF scheduler. A large value of t_c offers the opportunity of waiting a long time before scheduling a user when its channel quality hits a peak. We then expect the scheduler to better exploit multi-user diversity at the expense of longer packet delays. Thus the time constant should be set accounting for the packet delay tolerance of the applications [23]. For $t_c = 100$, the typical time-scale of the scheduler is around 100 ms, which is acceptable for most data applications.

Figure 1 illustrates the principle of the PF scheduler. The graph shows the feasible rate $r_1(t)$ of user 1, those slots where this user is selected, and the average throughput $T_1(t)$, for $n = 4$ and $t_c = 100$. Throughput values are normalized so that $R_1 = 1$. We assume the feasible data rate is linear in the SINR. The simulation results are obtained for Rayleigh fading with time correlations given by Jakes model with a Doppler shift corresponding to 4 slots [16]. We observe that the PF scheduler is indeed opportunistic in the sense that most selected slots correspond to local peaks of the feasible rate.

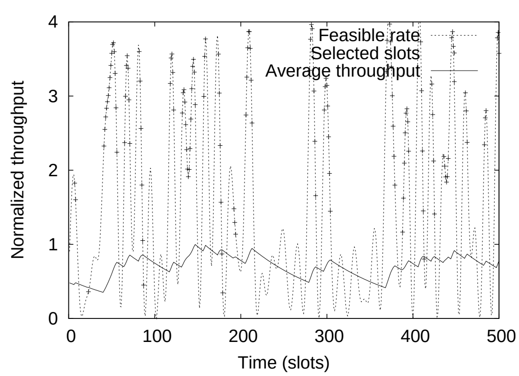

Figure 1: The PF scheduler: those slots with the best feasible rate to average throughput ratio are selected.

The ideal case. We first consider an ideal situation where users have symmetric fading statistics (i.e., $\xi_1(t), \dots, \xi_n(t)$ are i.i.d. random variables) and no data rate constraints (i.e., each user has an infinite backlog of data). In all the paper, we assume a perfect rate prediction (recall that hybrid ARQ may correct prediction errors). The PF scheduler is then asymptotically fair in the sense that all users receive the same fraction of the time slots [14]. In addition, when $t_c \to \infty$, the average throughputs are approximately constant and given by:

$$T_i = R_i \times \frac{G(n)}{n}, \tag{7}$$

where $G(n)$ denotes the scheduling gain, defined as the ratio of what a user receives compared to a blind round-robin scheduler. In view of (6) and (7), we have:

$$G(n) = \mathrm{E}[\max(\xi_1(t), \dots, \xi_n(t))]. \tag{8}$$

For Rayleigh fading and a linear data rate to SINR dependency for instance, $\xi_1(t), \dots, \xi_n(t)$ are exponentially distributed. We then obtain:

$$\begin{aligned} G(n) &= \int_0^\infty \Pr(\max(\xi_1(t), \dots, \xi_n(t)) > u)du \\ &= \int_0^\infty 1 - (1 - e^{-u})^n du \\ &= 1 + \frac{1}{2} + \dots + \frac{1}{n}. \end{aligned} \tag{9}$$

Thus the scheduling gain increases like $\ln(n)$ with n. Note that $G(n)/n \approx 0.5$ for $n = 4$, which indeed corresponds to the average normalized throughput observed in the example of Figure 1.

Impact of asymmetric fading. In practice, users do not experience the same fading. Fading is an extremely complex phenomenon caused by the interaction between the propagation environment and user mobility. While Rayleigh fading naturally arises from multi-path reflections, the presence of a significant line-of-sight component results in Rician fading [21]. The transmission data rate to SINR is also not linear, especially for high data rates, and depends on modulation and coding schemes. Thus, while typically mutually independent, $\xi_1(t), \dots, \xi_n(t)$ are generally not identically distributed.

Holtzman observed the bias of the PF scheduler against variable radio channels in asymmetric conditions [15]. Here we illustrate this result in the simple case of two users with $\xi_1(t) \sim \exp(1)$ and $\xi_2(t) = 1$. When $t_c \to \infty$, T_1 and T_2 are approximately constant so that user 2 is scheduled with probability:

$$p = \Pr(\xi_1 \frac{R_1}{T_1} < \frac{R_2}{T_2}).$$

Substituting the expressions for T_1 and T_2, we deduce $p \approx 2/3$. Thus user 2 is selected twice as often as user 1. Observing that the transmission rate of users far from the BS is generally more variable than that of close users (since the latter is typically close to the maximum data rate), we conclude that the PF scheduler favors close users and therefore tends to realize the maximum rate sharing considered in Section 2.

Impact of rate constraints. We now consider a situation where some users do not have an infinite backlog of data due to rate constraints. In practice, the data rate can indeed be limited by the wired network (e.g., the server) or by the mobile itself. Consider, for example, the widely used "stop-and-wait" error control protocol consisting in waiting for the acknowledgement of each packet before transmitting the next one [19]. This introduces a minimum delay between the transmission of successive packets (e.g., 4 slots for HDR systems [12]).

The impact of such rate constraints is that the average throughput of some users is much less than what the scheduler would normally allocate. In view of (6), these users are likely to be selected independently of their current transmission rate. Consider for instance the simple case of two users where user 1 has a rate constraint $T_1^{\max} << R_1$ while user 2 has no rate constraint. When $t_c \to \infty$, T_1 and T_2 are approximately constant so that when active, user 1 is selected with probability:

$$p = \Pr(\xi_1 \frac{R_1}{T_1} > \xi_2 \frac{R_2}{T_2}).$$

Using the fact that $T_1 \approx T_1^{\max}$, we deduce $p \approx R_1/(R_1 + T_1^{\max})$ for symmetric Rayleigh fading, so that user 1 is selected with a probability close to 1.

Figure 2 gives the simulation results obtained for the same scenario as that of Figure 1, except that users 2,3,4 have one packet to transmit every four slots (with $R_2 = R_3 = R_4 = 1$ packet/slot). We observe that user 1 only slightly benefits from the rate constraint of other users ($T_1 \approx 0.8$ instead of 0.5).

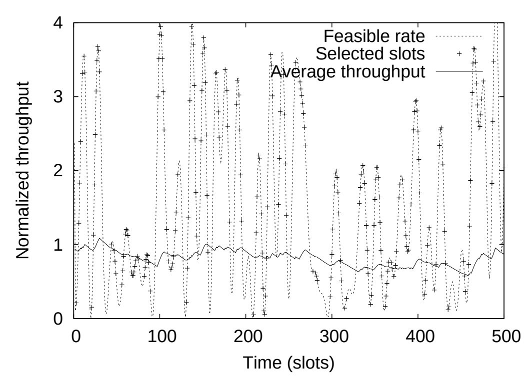

Figure 2: The PF scheduler with rate constraints.

4 Score-Based scheduler

We now present the principle and the key properties of a new algorithm referred to as the Score-Based (SB) scheduler, that behaves like the PF scheduler in the ideal case and whose performance does not suffer from asymmetric fading statistics or data rate constraints.

Principle. The SB scheduler selects at slot t the user $i(t)$ with the best score:

$$i(t) = \arg\min_{j=1,\dots,n} s_j(t)$$

where the score $s_j(t)$ of user j at slot t corresponds to the rank of its current transmission rate $r_j(t)$ among the past values $\{r_j(t), r_j(t-1), \dots, r_j(t-W+1)\}$ observed over a window of size W. If two users are active for instance and the current rate $r_1(t)$ of user 1 is in second position among its W past rate values while the current rate $r_2(t)$ of user 2 is in fifth position among its W past rate values, user 1 is selected at slot t.

Instead of selecting a user when its transmission rate is high relative to its own *average throughput* (the principle of the PF scheduler), the SB scheduler selects a user when its transmission rate is high relative to its own *rate statistics*. The corresponding time constant is given by the window size W, which should be set sufficiently large to track the distribution of the rate process while accounting for the packet delay tolerance of applications.

In case of equality of the current transmission rate with one or several past rate values, either order is chosen with equal probability. Formally, the score of user j at slot t is given by:

$$s_i(t) = 1 + \sum_{l=1}^{W-1} 1_{\{r_i(t) < r_i(t-l)\}} + \sum_{l=1}^{W-1} 1_{\{r_i(t) = r_i(t-l)\}} X_l,$$

where X_l are i.i.d. random variables on $\{0, 1\}$ with $\Pr(X_l = 0) = 1/2$. It is worth noting that the score may actually be evaluated based on any measure of the channel condition (e.g., SINR), not necessarily the transmission rate.

Figure 3 illustrates the principle of the SB scheduler, in exactly the same conditions as those of Figure 1, with $W = 100$. Again, we observe that the SB scheduler is opportunistic in the sense that most selected slots correspond to local peaks of the feasible rate.

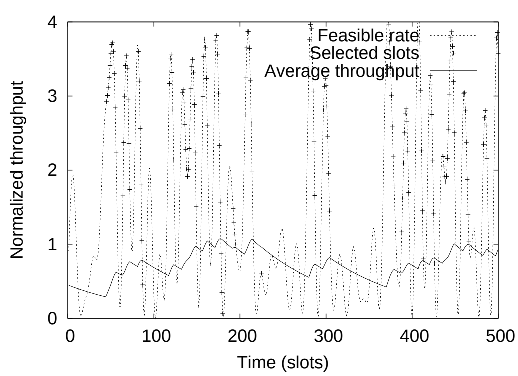

Figure 3: The SB scheduler: those slots with the best score are selected.

The ideal case. We first consider the ideal case described in Section 3. When $W \to \infty$, the SB scheduler captures the entire distribution of the rate process and the normalized score $s_j(t)/W$ of each user j is uniformly distributed over $[0, 1]$. In particular, the SB scheduler behaves exactly as the PF scheduler in the sense that all users receive the same fraction of the time slots and have a long-term average throughput given by (7) with the scheduling gain (8).

It turns out that the performance of the SB scheduler can also be evaluated for finite values of the window size. Provided W is sufficiently large so that the effect of time correlations becomes negligible, the score of each user is uniformly distributed over $\{1, \dots, W\}$. In particular, all users receive the same fraction of the time slots. We also deduce that for each $l = 0, \dots, W-1$:

$$\Pr(s_{i(t)} > l) = \Pr(\min_{j=1,\dots,n} s_j(t) > l) = \left(\frac{W-l}{W}\right)^n.$$

Now let $F(u) = \Pr(\xi_1(t) \le u)$ and $F_l(u)$ be the probability that among W values of the rate process, the l highest values are larger than u and the other values are less than u:

$$F_l(u) = \binom{W}{l} F(u)^{W-l}(1 - F(u))^l.$$

We have:

$$\begin{aligned}
\Pr(\xi_{i(t)}(t) \le u) &= \sum_{k=1}^{W} \sum_{l=0}^{k-1} F_l(u) \Pr(s_{i(t)} = k) \\
&= \sum_{l=0}^{W-1} \Pr(s_{i(t)} > l) F_l(u) \\
&= \sum_{l=0}^{W-1} \left(\frac{W-l}{W}\right)^n F_l(u).
\end{aligned}$$

Using the fact that $\sum_{l=0}^{W} F_l(u) = 1$, we deduce the scheduling gain:

$$G(n) = \sum_{l=1}^{W} \left[1 - \left(\frac{W-l}{W}\right)^n\right] \int_0^\infty F_l(u) du. \quad (10)$$

For Rayleigh fading, we get:

$$G(n) = \sum_{l=1}^{W} \left[1 - \left(\frac{W-l}{W}\right)^n\right] \frac{1}{l}. \quad (11)$$

When $W \to \infty$, this expression tends to the optimal gain (9). Figure 4 gives the corresponding scheduling gain. We observe that the gain increases with the window size and is almost optimal when $W = 100$ for typical values of the number of active users (in view of (3), n exceeds 10 with probability less than 0.1 for $\rho = 0.8$).

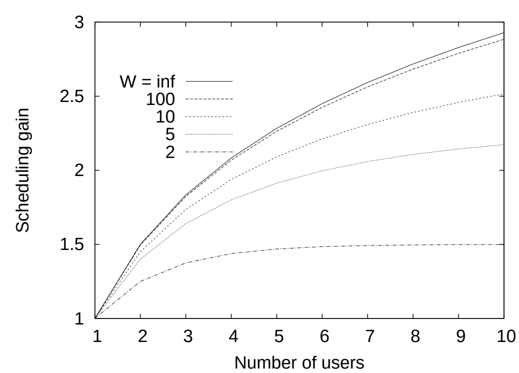

Figure 4: Impact of the window size W on the scheduling gain.

Impact of asymmetric fading. A key property of the SB scheduler is that the results derived in the ideal case apply equally in asymmetric fading conditions. This is because the score of each user is uniformly distributed over $\{1, \ldots, W\}$, independently of the fading process of other users. In particular, the SB scheduler is perfectly fair at slot level and the throughput received by a given user is the same *as if* all users had the same fading process. We deduce the long-term throughput of user i:

$$T_i = R_i \times \frac{G_i(n)}{n},$$

where $G_i(n)$ is the scheduling gain associated with the fading process of user i, i.e., that given by (10) with functions F_l corresponding to the distribution function $F(u) = \Pr(\xi_i(t) \leq u)$. Thus the scheduling gain of each user depends on the number of active users and its *own* rate statistics only.

Impact of rate constraints. As the SB scheduler is not based on any measure of average throughput, it does not suffer from rate constraints like the PF scheduler. Figure 5 is the analog of Figure 2 for the SB scheduler with a window size $W = 100$. We observe that user 1 strongly benefits from the rate constraint of other users ($T_1 \approx 1.1$ instead of 0.5). The impact of rate constraints on the PF scheduler and the SB scheduler is further analysed in the next section.

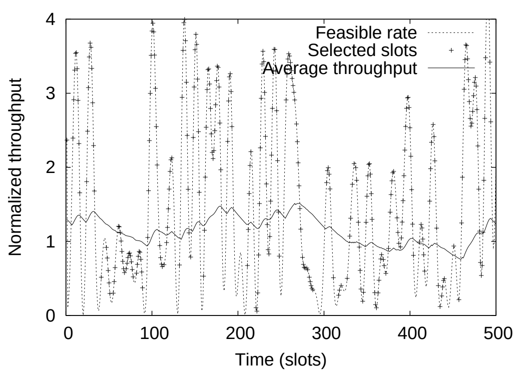

Figure 5: The SB scheduler with rate constraints.

5 Comparison by simulation

In this section, we compare the performance of the PF scheduler and the SB scheduler by means of simulation. The respective time constants are $t_c = 100$ and $W = 100$. Each point of the following graphs correspond to an average over 100 independent simulation runs of 36000 slots each (corresponding to 1' for HDR systems, 24" for HSDPA systems).

The ideal case. Figure 6 gives the scheduling gains of the PF scheduler and the SB scheduler for Rayleigh fading in the absence of rate constraint. The simulation results confirm that both schedulers have the same performance in the ideal case, given by the theoretical gain (11) with $W = 100$.

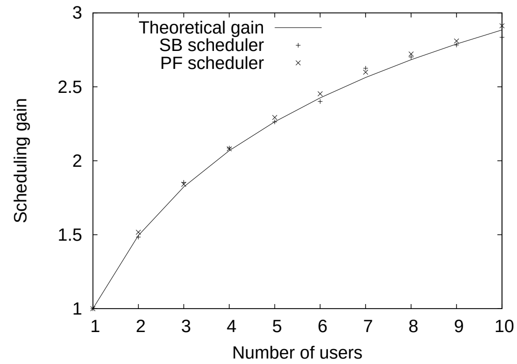

Figure 6: Comparison of PF and SB in the ideal case.

Impact of asymmetric fading. Figure 7 gives the scheduling gains for user 1 (upper points) and other users (lower points) when user 1 has Rayleigh fading while other users have no fading. We observe that, while the SB scheduler behaves for each user exactly as in symmetric fading conditions, the PF scheduler is biased against user 1 due to the more variable rate statistics of this user (cf. Section 3).

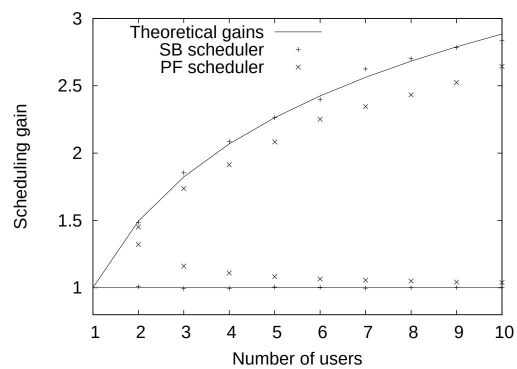

Figure 7: Comparison of PF and SB in asymmetric fading conditions.

Impact of rate constraints. Figure 8 gives the scheduling gains for user 1 when user 1 has no rate constraint while the other users have one packet to transmit every n slots (with $R_2 = \ldots = R_n = 1$ packet/slot), assuming symmetric Rayleigh fading. We observe that the SB scheduler is much more efficient than the PF scheduler. This is due to the fact that the PF scheduler tends to select users with rate constraints independently of their current feasible rate (cf. Section 3). The SB scheduler, on the other hand, always selects the user with the best score and therefore better exploits multi-user diversity.

6 Conclusion

We have first demonstrated by the analysis of flow-level dynamics that sharing the time slots in a fair way is efficient in terms of average user performance. We then have shown that, while fair and indeed opportunistic in the ideal case, the PF scheduler may be unfair and unable to fully exploit multi-user diversity in more realistic cases. Finally, we have presented the principle and

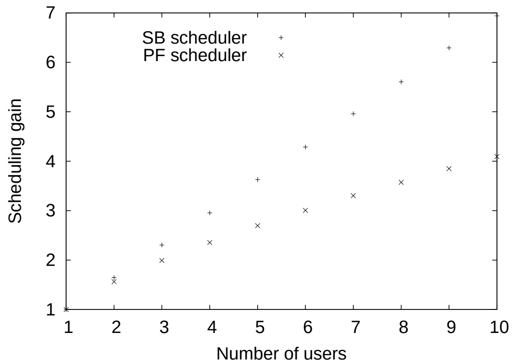

Figure 8: Comparison of PF and SB in the presence of rate constraints.

the key properties of a new algorithm referred to as the Score-Based (SB) scheduler. Instead of selecting a user when its transmission rate is high relative to its own *average throughput*, the SB scheduler selects a user when its transmission rate is high relative to its own *rate statistics*. At the expense of a slightly higher complexity, this scheduler behaves like the PF scheduler in the ideal case but its performance does not suffer from asymmetric fading statistics or data rate constraints.

REFERENCES

[1] 3GPP TS 25.214, "Physical Layer Procedures (FDD)". Release 5, 2003.

[2] 3GPP TS 25.308, "UTRA High Speed Downlink Packet Access (HSDPA); Overall description". Release 5, 2003.

[3] 3GPP2 C.S0024, "cdma2000 High Rate Packet Data Air Interface Specification". 2002.

[4] R. Agrawal, V. Subramanian, "Optimality of certain channel-aware scheduling policies". Proc. 40th Annual Allerton Conf. Commun. Control Comp., 2002.

[5] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushayana and A. Viterbi, "CDMA/HDR: A bandwidth-efficient high-speed wireless data service for nomadic users". IEEE Commun. Magazine, 70–77, July 2000.

[6] F. Berggren and R. Jäntti, "Asymptotically fair scheduling in fading channels". Proc. of IEEE VTC Fall, 2002.

[7] T. Bonald and L. Massoulié, "Impact of fairness on Internet performance". Proc. of ACM SIGMETRICS, 2001.

[8] T. Bonald and A. Proutière, "Wireless downlink channels: user performance and cell dimensioning". Proc. of ACM MOBICOM, 2003.

[9] S. Borst, "User-level performance of channel-aware scheduling algorithms in wireless data networks". Proc. of IEEE INFOCOM, 2003.

[10] S. Borst, P. Whiting, "Dynamic channel-sensitive scheduling algorithms for wireless data throughput optimization". IEEE Trans. Veh. Techn. (52) 569–586, 2003.

[11] E.F. Chaponniere, P.J. Black, J.M. Holtzman, D.N.C. Tse, "Transmitter directed code division multiple access system using path diversity to equitably maximize throughput". US Patent 6449490, 2002.

[12] E. Esteves, P.J. Black and M.I. Gurelli, "Link adaptation techniques for high-speed packet data in third generation cellular systems". Proc. of European Wireless Conference, 2002.

[13] G. Fayolle, I. Mitrani and R. Iasnogorodski, "Sharing a processor among many classes". Journal of the ACM (27) 519–532, 1980.

[14] J.M. Holtzman, "CDMA forward link waterfilling power control". Proc. of IEEE VTC Spring, 2000.

[15] J.M. Holtzman, "Asymptotic analysis of Proportional Fair algorithm". Proc. of 12th IEEE PIMRC, 2001.

[16] W.C. Jakes, (ed.), "*Microwaves mobile communications*". Wiley, New York, 1974.

[17] H.J. Kushner, P.A. Whiting, "Asymptotic properties of Proportional Fair sharing algorithms". Proc. 40th Annual Allerton Conf. Commun. Control Comp., 2002.

[18] X. Liu, E. K.P. Chong, and N.B. Shroff, "A framework for opportunistic scheduling in wireless networks". Computer Networks (41-4) 451–474, 2003.

[19] S. Parkvall, E. Dahlman, P. Frenger, P. Beming, M. Persson, "The high speed packet data evolution of WCDMA". Proc. of the 12th IEEE PIMRC, 2001.

[20] S. Shakkottai, A.L. Stolyar, "Scheduling algorithms for a mixture of real-time and non-real time data in HDR". Proc. of ITC 17, 2001.

[21] B. Sklar, "Rayleigh fading channels in mobile digital communication systems". IEEE Commun. Magazine, 90–100, July 1997.

[22] V. Tsibonis, L. Georgiadis, L. Tassiulas, "Exploiting wireless channel state information for throughput maximization". Proc. of IEEE INFOCOM, 2003.

[23] P. Viswanath, D. Tse and R. Laroia, "Opportunistic beamforming using dumb antennas". IEEE Trans. on Information Theory (48) 1277–1294, 2002.